

VOLUME 1

2011 INTERNATIONAL CONFERENCE ON

# UNCERTAINTY REASONING AND KNOWLEDGE ENGINEERING

4-7 AUGUST 2011  
BALI, INDONESIA



# Proceedings of the International Conference on Uncertainty Reasoning and Knowledge Engineering

4-7 August 2011, Bali, Indonesia

## **URKE 2011**

---

4  
8  
2  
6  
0  
4  
8

All right reserved.

**Print Version**

*Copyright and Reprint Permission:* Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Operations Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved. Copyright 2011 by the Institute of Electrical and Electronics Engineers



URKE 2011  
Table of Contents  
Volume – 1

Preface – Volume 1.....	VI
Conference Organization – Volume 1 .....	VII
Alternative implementation techniques of parallelized differential evolution for multi-core processors.....	1
<i>Kiyaharu Tagawa and Takashi Ishimizu</i>	
Designing of expert system for troubleshooting diagnosis on gas chromatography gc-2010 by means of inference method.....	5
<i>Suyanto M. Abduh Pancaputra, and Ruri Agung Wahyuono</i>	
Selection of districts for data entry sites in aceh household census.....	9
<i>Bonivasius Prasetya, Stefanie Intan Christienova, and Idaman</i>	
Collaborative internet content filtering on the internet infrastructure in Malaysia.....	12
<i>Fajri Achmad Maulana, Sahidan Abdulmana, and Fauzan Alfariti</i>	
Reinforcement mechanism and engineering applications of jointed rock mass.....	16
<i>Jingsui Zhang, Fei Wang, Shaojun Fu, and Tingting Liao</i>	
Uncertainty, protocol behavior prediction through chaos theory?.....	20
<i>Prabhu Shankar Kaliappan</i>	
Soaking with uncertainty for problem solving.....	24
<i>Wei Hui and Fu Xixu</i>	
A model for information systems evaluation.....	28
<i>Liang-Chuan Wu</i>	
Flexible decision making in uncertain R&D projects.....	32
<i>Juite Wang</i>	
Semantic technology classification - a defence and security case study.....	36
<i>Dirk Thorleuchter and Dirk Van den Poel</i>	
Companies website optimising concerning consumer's searching for new products.....	40
<i>Dirk Thorleuchter and Dirk Van den Poel</i>	
A framework for vision-based swimmer tracking.....	44
<i>Wen-Hui Chen, Po-Chuan Cho, Ping-Lin Fan, and Yi-Wen Yang</i>	
Initiating layers architecture design for software product line.....	48
<i>I Made Murwantara</i>	

Research on remote intelligent monitoring system.....	52
<i>Wang Xiao-li, Gao Xiao-hong, and Li Jian-wei</i>	
Empirical study on selection and evaluation of TPL based on CS.....	56
<i>Wei Xu, Songzheng Zhao, and Liangdong Lu</i>	
Research on the application of neural network PID to wipe film motor of molecular distillation.....	60
<i>Guan Changjun and Li Hui</i>	
Design of a pet drinking vessel based on TRIZ & growth design.....	64
<i>Hongwu Chen, Xuan Liu, and Can Wang</i>	
Knowledge sharing platform for project team based on web feeds.....	67
<i>I-Ching Hsu, Lee Jang Yang, and Der-Chen Huang</i>	
New switched filtering method for recurrent neural networks.....	71
<i>Choon Ki Ahn</i>	
Research on China's studying abroad situation and countermeasures.....	75
<i>Chen Aijuan and Zhang Qian</i>	
Classifier combination for telegraphese restoration.....	79
<i>Leo Willyanto Santoso</i>	
Design and implementation of the soft-sensing system for fermentation process.....	83
<i>Wu Zi-yue, Xu Zhe, and Jinfeng Geng</i>	
The study of mobile education development based on 3G technique and cloud computing.....	86
<i>Huang Bin</i>	
The analysis of human judgment accuracy using decision tree models.....	90
<i>Hsi-Peng Lu, Yi-Wen Yang, and Wen-Hui Chen</i>	
A weighted fuzzy time series model based on adoptive OWA operators.....	94
<i>Jing-Rong Chang, and Yu-Jie Huang</i>	
Research on randomized greedy algorithm for k-median problem.....	98
<i>Wang Shouqiang</i>	
An algorithm for splitting an orthogonal polyhedron with a polyplane.....	102
<i>Jefri Marzal, Hong Xie, and Chun Che Fung</i>	
Implementing user-oriented interfaces: from user analysis to framework's components.....	107
<i>Sergio Inzunza, Andrés Mejía, Reyes Juárez-Ramírez, and Manuel Gómez-Ruelas</i>	
Towards a formal model for knowledge transference:	
An initial proposal to support the adoption of process models.....	111
<i>Violeta ocegueda-miramontes, Reyes Juárez-Ramírez, Rafael Pimienta-Romo, and Manuel Gómez-Ruelas</i>	
A formal approach for measuring the lexical ambiguity degree in	
Natural language requirement specification.....	115
<i>Carlos Huertas, Reyes Juárez-Ramírez, Manuel Gómez-Ruelas, and Héctor Plata</i>	
Towards improving user interfaces: a proposal for integrating functionality	
and usability since early phases.....	119



<i>Reyes Juárez-Ramírez, Alan A. Gutiérrez, Manuel Gómez-Ruelas, and Pavel Negrete</i>	
Using iterative learning control methods for 2-D systems.....	124
<i>Zahra Maghsoudzade Sarvestani, Ahmadreza Argha, and Mehdi Roopaei</i>	
Direct torque control of brushless dc motor drives with reduced starting current using fuzzy logic controller.....	129
<i>N. Parhizkar, M. Shafiei, and M. Bahrami Kouhshahi</i>	
Application of topic based vector space model with wordnet.....	133 ←
<i>Adi Wibowo and Andreas Handojo</i>	
Processing comparisons and evaluations in business intelligence: a question answering system.....	137
<i>Kenston Choi, Rosalyn Margret Pacaña, Adrian Lester Tan, Jonathan Yiu, and Nathalie Rose Lim</i>	
A study of applying arima and svm model to software reliability prediction.....	141
<i>Jung-Hua Lo</i>	
Incorporating information quality criteria in blog search engine.....	145
<i>Fatemeh Azimzadeh, Abd. R. Ramli, Borhanoddin. M. Ali, and Hamidah Ibrahim</i>	
Study on the development trend of the digital printing.....	149
<i>Qian Hua</i>	
Using particle swarm optimization algorithm for image enhancement.....	154
<i>Hanxing Chen and Jun Tian</i>	
Chaos-based gray image watermarking algorithm.....	158
<i>Junmei Guo, Yong Yang, and Ning Wang</i>	
Special equipment safety analysis and countermeasures.....	161
<i>Jun Tian, and Hanxing Chen</i>	
A study on home information network.....	165
<i>Yong Yang and Junmei Guo</i>	
Non-subsampled contourlets and gray level co-occurrence matrix based images segmentation.....	168
<i>Zhang Jian and Chen Xiaowei</i>	
Preparation for strustural path analysis from plain text input output matrix.....	171
<i>Nova Nurviana, Rizchi Wahyuni, and Idaman</i>	
A study on the design of half pipe.....	175
<i>Zhao Yang and Li zhaonan</i>	
Evaluation of building materials logistics provider based on FAHP—TOPSIS.....	180
<i>Sun Shusheng, Zhang Qin, and Zhang Min</i>	
An approach for multimodal biometric fusion under the missing data scenario.....	185
<i>Quang Duc Tran, Panos Liatsis, Bing Zhu, and Changzheng He</i>	
Hybrid-dimension association rules for diseases track record analysis at Dr. Soetomo general hospital.....	189
<i>Silvia Rostianingsih, Gregorius Satia Budhi, and Ni Wayan Yessy Dwijayanti</i>	
The forecast for the engine wear trend based on grey gm(1,1) and markov chain model.....	192
<i>Zhang Yong, Chen Liang, and Gu Xiaoxing</i>	

Data mining market basket analysis' using hybrid-dimension association rules, case study in minimarket X.....	196
<i>Djoni Haryadi Setiabudi, Gregorius Satia Budhi, I Wayan Jatu Purnama, and Agustinus Noertjahyana</i>	
Product cost estimation using supper efficiency data envelopment analysis.....	200
<i>Fantahun M. Defersha, Adil Salam, and Nadia Bhuiyan</i>	
Prioritization of information quality criteria in the blog context.....	205
<i>Mohammad Javad Kargar</i>	
Towrads dedicated information quality frameworks in social network products.....	209
<i>Mohammad Javad Kargar</i>	
Study on the fast image registration based on genetic algorithm.....	213
<i>Shuai Feng, Yufeng Li, and Weiping Zhao</i>	
Compression and realization of power quality disturbance data based on wavelet analysis.....	217
<i>Zhao Hongtu and Xi Dongmei</i>	
Research on face detection algorithm in instant message robot.....	220
<i>Yufeng Li, Haijuan Zhang, and Yanan Zhang</i>	
Three-dimensional visual geographic information system.....	224
<i>Bo Hang</i>	
Runtime monitoring web services implemented in BPEL.....	228
<i>Zhang Haiteng, Shao Zhiqing, and Zheng Hong</i>	
Based on stm32f103 implement profibus-dp slave with high-speed transmission.....	232
<i>Xu Hongyan, Wu Guichu, and Chen Chong</i>	
LVMM: a lightweight virtual machine memory management Architecture for virtual computing environment.....	235
<i>Weizhe Zhang, Tao Cheng, Hui He, and Albert M.K. Cheng</i>	
Primary signal detection over rayleigh fading channel for cognitive radio.....	239
<i>Yufeng Li, Yanan Zhang, and Haijuan Zhang</i>	
A spt-routing key management scheme for heterogeneous wireless sensor networks based on ECC.....	243
<i>Gang Han, Rui Zhou, and Hua Yang</i>	
Design of video surveillance system based on 3G wireless network in underground coal mine.....	248
<i>Wei Zhou</i>	
A hybrid key management scheme for heterogeneous wireless sensor networks based on ecc and trivariate symmetric polynomial.....	251
<i>Rui Zhou and Hua Yang</i>	
Comprehensive evaluation and selection model of supplier under E-business.....	256
<i>Wang Mian</i>	

Voice conversion application (VOCAL).....	259
<i>Liliana, Resmana Lim, and Elizabeth Kwan</i>	
Clustering search engine at petra christian university library using suffix tree clustering.....	263 ←
<i>Andreas Handojo, Adi Wibowo, and Jemmy Lay Santo</i>	
Author Index – Volume 1 .....	267



**IEEE Conference Publications Management Group**  
**445 Hoes Lane**  
**P.O. Box 1331**  
**Piscataway, NJ 08854**

IEEE Catalog Number: CFP1115N-PRT  
ISBN: 978-1-4244-9983-0



## Application of Topic Based Vector Space Model with WordNet

Adi Wibowo  
Informatics Department  
Petra Christian University  
Surabaya, Indonesia  
adiw@peter.petra.ac.id

Andreas Handojo  
Electrical Department  
Petra Christian University  
Surabaya, Indonesia  
handojo@peter.petra.ac.id

Albert Halim  
Informatics Department  
Petra Christian University  
Surabaya, Indonesia

**Abstract**—Topic Based Vector Space Model (TVSM) proposed a new vector space that its dimensions is composed of topics. Every term and document is represented by vectors inside this vector space. By using topics as dimensions TVSM tries to overcome word-mismatch between terms with similar topics in finding relevant documents to query. This study proposes to develop relations between terms using WordNet and thesaurus to help TVSM calculating similarity between documents. Relations between terms are represented by relation score. This study proposes a way to find optimal relation score for a set of documents. To help indexing documents with multi language terms this study also proposes to use dictionary to expand query terms.

*Topic based vector space model, wordnet, dictionary*

### I. INTRODUCTION

Search engine is designed to find documents relevant to user queries. One of a method used in measuring similarity is Vector Space Model (VSM). VSM assumes every unique term from documents can be represented by each dimension of vector [1]. This method also has the assumption that each term in the document are independent. A term “lion” has no relationship or similarity with “animal”, or “car” with “automotive”. This condition is called word mismatch. Word mismatch is one of factor that causes a decrease in recall and lead to less accurate document similarity calculation process.

To overcome the problem Topic Based Vector Space Model (TVSM) has the assumption that each term has a relationship with other terms. Their proximity can be measured by using relationship from thesaurus [2]. Term generalization can also use dictionary, so it is possible to search multi language document and still provide relevant documents to user queries.

This paper tries to propose the application of WordNet as thesaurus and english-indonesian dictionary to provide terms and relationships needed by TVSM.

### II. VECTOR SPACE MODEL AND TVSM

In Vector Space Model every document is represented by a vector in vector space. Each vector dimension represents every unique term found in all documents. Similarity between documents is calculated by measuring angle between their vector. In Vector Space Model documents and query are represented as equation 1.

$$\begin{aligned} d_i &= (w_{1,i}, w_{2,i}, w_{3,i}, \dots, w_{t,i}) \\ q &= (w_{1,q}, w_{2,q}, w_{3,q}, \dots, w_{t,q}) \end{aligned} \quad (1)$$

Similarity measurement between document and query is performed by using cosine similarity measure.

$$Sim(q, d_i) = \frac{\sum_i w_{q,j} w_{i,j}}{\sqrt{\sum w_{q,j}^2} \sqrt{\sum w_{i,j}^2}} \quad (2)$$

$w_{q,j}$  is a weight of term  $j$  from query  $q$ , and  $w_{i,j}$  is a weight of term  $j$  from document  $i$ . Each weight is calculated from  $tf \cdot idf$  value.

$$w_{i,j} = tf_{i,j} * \log \left( \frac{D}{df_j} \right) \quad (3)$$

$tf_{i,j}$  is a number of term  $j$  in a document  $i$ ,  $df_j$  is a number of documents that contain term  $j$ , and  $D$  is total number of documents.

Topic Based Vector Space Model assumes another form of vector. If the vector dimension of VSM represents terms, the vector dimension of TVSM represents fundamental topic of terms [2]. Terms in TVSM is also represented by the vector. There is  $d$  dimensional space  $R$  which only has positive axis-intercepts.

$$R \in \mathbf{R}_{\geq 0}^d \text{ with } d \in \mathbf{N}_{>0} \quad (4)$$

Each term  $i \in \{1, \dots, n\}$  is represented by term vector  $\vec{t}_i$  and has term weight between 0 and 1. Term weight in TVSM is defined as:

$$\vec{t}_i = (t_{i1}, t_{i2}, \dots, t_{id}) \in R \quad (5)$$

$$|\vec{t}_i| = \sqrt{t_{i1}^2 + t_{i2}^2 + \dots + t_{id}^2} \in [0; 1] \quad (6)$$

And term correlation is defined as :

$$\overrightarrow{t_i t_j} = |\overrightarrow{t_i}| \cdot |\overrightarrow{t_j}| \cdot \cos(\omega_{ij}) \quad (7)$$

$$i, j \in \{1, \dots, n\}$$

$$i : |\overrightarrow{t_i}| \in [0; 1]$$

$$\omega_{ij} \in [0^\circ; 90^\circ]$$

A document  $k \in \{1, \dots, m\}$  is represented by vector  $\overrightarrow{d_k} \in R$ , where vector length is standardized to one.

$$\overrightarrow{d_k} = \frac{1}{\delta_k} \tilde{\delta}_k \rightarrow |\overrightarrow{d_k}| = 1$$

$$\tilde{\delta}_k = \sum_{i=1}^n e_{ki} \overrightarrow{t_i} \quad (8)$$

while

$$e_{ki} = \text{number of term } i \text{ in document } k$$

With this document representation, similarity between document  $k$  and  $l$  is defined as scalar product from vector documents.

$$\overrightarrow{d_k} \cdot \overrightarrow{d_l} = |\overrightarrow{d_k}| \cdot |\overrightarrow{d_l}| \cdot \cos(\omega_{kl}) = \cos(\omega_{kl})$$

$$\text{where } \overrightarrow{d_k} = \overrightarrow{d_l} = 1$$

$$\overrightarrow{d_k} \cdot \overrightarrow{d_l} = \frac{1}{|\delta_k|} \tilde{\delta}_k \cdot \frac{1}{|\delta_l|} \tilde{\delta}_l = \frac{1}{|\delta_k| |\delta_l|} \sum_{i=1}^n \sum_{j=1}^n e_{ki} e_{lj} \overrightarrow{t_i} \cdot \overrightarrow{t_j} \quad (9)$$

The length of document vector  $\overrightarrow{d_k}$  is defined as:

$$\begin{aligned} |\delta_k| &= \left| \sum_{i=1}^n e_{ki} \overrightarrow{t_i} \right| = \sqrt{\left( \sum_{i=1}^n e_{ki} \overrightarrow{t_i} \right)^2} \\ &= \sqrt{\sum_{i=1}^n \sum_{j=1}^n e_{ki} e_{kj} \overrightarrow{t_i} \cdot \overrightarrow{t_j}} \end{aligned} \quad (10)$$

### III. WORDNET

WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. WordNet provides a network of meaningfully related words and concepts [3].

Synsets are term lists or collocation that have the same meaning and their use in particular context are interchangeable. Relations between terms used to describe relation between synsets are:

1. **Hypernym.** A word whose meaning denotes a superordinate or superclass. Animal is a hypernym of dog.
2. **Hyponym.** A word whose meaning is included in that of another word. Hyponym is an opposite of hypernym.

3. **Holonym.** A word for the whole of which other words are part. Building is a holonym of window.
4. **Meronym.** A word that names a part of a larger whole. Window is a meronym of building.
5. **Coordinate term.** A word is a coordinate term of other word if both words share similar hypernym. Wolf is a coordinate term of dog, and dog is a coordinate term of wolf.

Example of WordNet synsets are shown at table 1.

TABLE I. WORDNET SYNSET EXAMPLES

Terms Related to Marketing	
Type of Relation	Related Words
Synonym	commerce, commercialism, mercantilism
Hypernym	commerce, commercialism, mercantilism, shopping
Hyponym	bait and switch, private treaty, dumping, distribution channel, channel
Holonym	marketing
Meronym	promotion, publicity, promotional material, packaging,
Coordinate term	commerce, commercialism, mercantilism, shopping

### IV. IMPLEMENTATION

This study proposed to provide terms relations to TVSM by using WordNet from Princenton University, and Indonesian Thesaurus from Indonesian National Education Department Language Center. To create relationship between the two languages, English and Indonesian, this study used Stardict Dictionary.

#### A. Preparation

Every term found from documents will be checked to eliminate stop words. Stop words is a list of common or general terms (e.g., prepositions, and articles) that are not significant because they appear in too many records. Examples of stop words in English are 'a', 'the', 'an', 'for', 'of', etc. Example of stop words in Indonesian Language are 'di', 'ke', 'dari', 'bahwa', 'pada', etc. To remove stop words this study used English stop word list from Gerard Salton and Chris Buckley [4]. As for Indonesian Language, this study used a list from Indonesian Grammar from Moeliono [5].

The resulting terms will be stemmed using Porter Stemmer for English terms [6], or Indonesian stemming algorithm from Nasief and Adriani for Indonesian terms. For English terms before processed by stemmer, they will be checked if they are irregular verbs. Irregular verbs have to be restored to their base form.

#### B. Developing Relations Between Terms

WordNet synsets and Indonesian Thesaurus are used to create relations between terms. A relation is a score that represents proximity between terms. An optimum score will be determined after several testing. To develop relations between terms, several condition applied:

- Stardict is used to create relationship between English terms and Indonesian terms.



- The score of relation is assumed to be between 0 and 1.
- A term is always a word.
- If a term consists of one word is related to other terms consists of multiple words then the score of relation is divided evenly between each word.

When a query is entered into system several steps are performed:

- If a query is consists of more than one word then the query is separated into multiple terms.
- Every query term will go through preparation steps first.
- By using Stardict similar words from other language will be retrieved. This set of words will be called expanded terms. The query now consists of expanded terms.
- Related terms to each term from a query is determined by using synsets from WordNet and Indonesian Thesaurus. The relation score of each interrelated terms is already predefined using tests at chapter IV.C. For every other terms that has no relationship with terms from query will have relation score equal to zero.
- By using related terms and relation score TVSM is performed to find relevant documents to a query.

### C. Finding Optimum Relation Score

In TVSM,  $e_{ki}$  is defined as the number of term  $i$  in document  $k$ .  $e_{ki}$  is used to calculate the weight of document  $k$  and similarity between documents. This study proposed to use of weighted  $e_{ki}$  by multiplying  $e_{ki}$  with inverse document frequency (IDF). IDF measures the importance of a term in a document.

$$idf_t = \log \frac{|D|}{|\{d \in D | t \in d\}|} \quad (11)$$

where

$$\begin{aligned} idf_t &: IDF \text{ for term } t \\ |D| &: \text{the total number of documents} \\ |\{d \in D | t \in d\}| &: \text{number of documents where the term } t \text{ appears} \end{aligned}$$

The method used in these tests is by searching a set of documents and measures Rank Weighted Average (RWA) scores. The assumption used by RWA is the more relevant the document it should be placed at the top rank. The lower score of RWA will provide better search engine results.

$$RWA = \text{average} \left( \frac{\text{document rank in search result}}{\text{document rank in relevant assessment}} \right) \quad (12)$$

The tests conducted to 350 documents with nine different relation score candidates between 0.1 and 0.9 with 0.1 interval to find which number provide better results. Each relation score is represented by R1 to R9 respectively. The queries used by the test is shown at table 2.

TABLE II. QUERIES USED BY TESTS

ID Query	Query terms
Q1	memilih mobil idaman
Q2	merapi mountain eruption
Q3	desain arsitektur rumah
Q4	supply chain management
Q5	data minings

The result of each test is shown at figure 1 to figure 5.

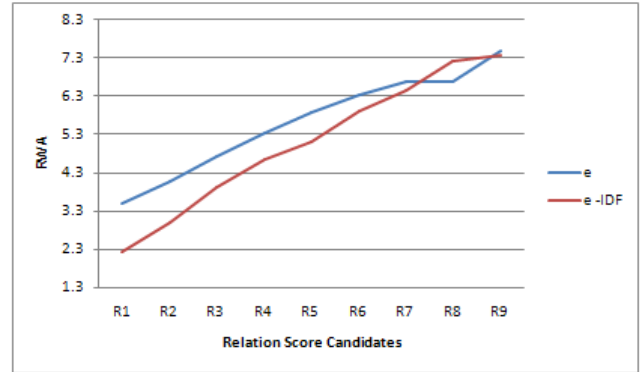


Figure 1. RWA from testing using query 1.



Figure 2. RWA from testing using query 2.



Figure 3. RWA from testing using query 3.

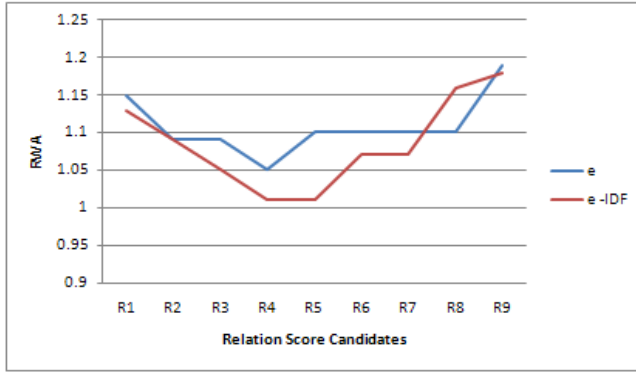


Figure 4. RWA from testing using query 4.

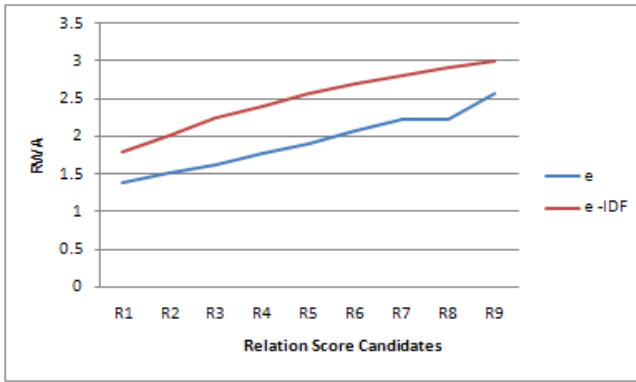


Figure 5. RWA from testing using query 5.

The average score of RWA from five tests shown at table III and figure 6 shows that e-IDF can help TVSM to obtain better RWA scores. These tests also shows that relation score R1 (0.1) provides better RWA compared with other relation scores.

	<i>e</i>	<i>e-IDF</i>
<b>R1</b>	1.929623	1.7516
<b>R2</b>	2.079487	1.944
<b>R3</b>	2.246713	2.1768
<b>R4</b>	2.510113	2.3986
<b>R5</b>	2.657908	2.5666
<b>R6</b>	2.881082	2.8124
<b>R7</b>	2.996781	2.97
<b>R8</b>	2.99821	3.2138
<b>R9</b>	3.280442	3.2702
<b>MEAN</b>	<b>2.62004</b>	<b>2.567111</b>

TABLE III. AVERAGE SCORES OF RWA FROM FIVE TEST.

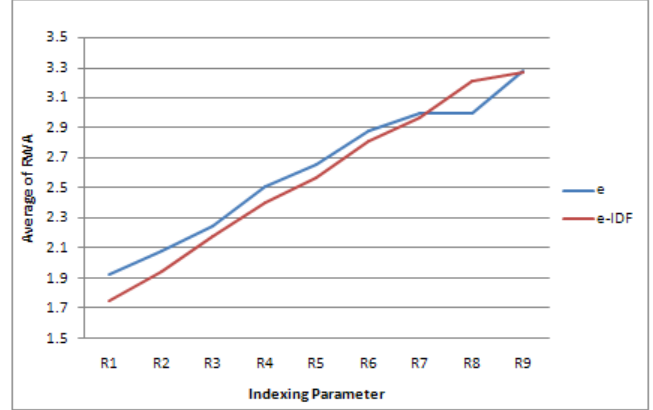


Figure 6. The average of RWA from five test for each relation score candidate.

## V. CONCLUSION

This study proposes the use of WordNet and Indonesian thesaurus to build relations between terms and also propose a way to provide relation scores between terms. This study also proposes the use of dictionary to bridge relations between terms from different languages.

## REFERENCES

- [1] G. Salton, Automatic Information Organization and Retrieval, 1968
- [2] J. Becker, D. Kuropka. "Topic based vector space model." USA: Business Information System, Proceedings of BIS, Colorado Springs. 2003
- [3] Princeton University "About WordNet." WordNet. Princeton University. 2010. <http://wordnet.princeton.edu>
- [4] C. Buckley dan G. Salton, Stopword List, Cornell University
- [5] A.M.Moeliono, et.al. "Indonesian Grammar". Balai Pustaka: Department of Education and Cultures, 1988.
- [6] M. Porter, "The Porter Stemming Algorithm". Retrieved May 24, 2011 from <http://snowball.tartarus.org/algorithms/porter/stemmer.html>.